

# **Historic, Archive Document**

Do not assume content reflects current scientific knowledge, policies, or practices.



aSDII  
A42  
Reserve

Cat / Dm / 68



United States  
Department of  
Agriculture

Forest Service

Rocky Mountain  
Forest and Range  
Experiment Station

Fort Collins,  
Colorado 80526

General Technical  
Report RM-GTR-265

# Selecting Stands for the Forest Planning Data Base: Sampling Background and Application

Dennis M. Donnelly and Phil R. Krueger



## **Abstract**

Donnelly, Dennis M.; Krueger, Phil R. 1994. Selecting stands for the forest planning data base: sampling background and application. Gen. Tech. Report RM-GTR-265. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 27 p.

The data base selected by a National Forest planning team for its Plan revision work must represent all the conditions that are part of the planning situation. This report describes a process to select sample stands necessary for planning from a forest vegetation database. We summarize basic sampling principles, use short examples to illustrate key points, and demonstrate actual practice through application on the Medicine Bow National Forest.

**Keywords:** Forest planning, planning, sampling, Medicine Bow National Forest, forest stands, data base

## **The Authors**

**Dennis M. Donnelly** is an Operations Research Analyst/Forester with the Forest Management Service Center (Washington Office Detached), Fort Collins, Colorado. He holds a Ph.D. in Forestry from Colorado State University. Dennis develops forest growth models, applies forest growth modeling research to new and current local variants of the Forest Vegetation Simulator, and consults with FVS users about its application in forest planning and silvicultural analysis. He has been with the Forest Service since 1971.

**Phil R. Krueger** is a Forester with the Supervisor's Office of the Medicine Bow/Routt National Forest in Laramie, Wyoming. He graduated from the University of Illinois with a B.S. in Forest Management. Phil provides FVS consultation and silviculture expertise to planning and projects for the Medicine Bow and Routt National Forests. He has been with the Forest Service since 1973.

# Selecting Stands for the Forest Planning Data Base: Sampling Background and Application

Dennis M. Donnelly, Operations Research Analyst  
Forest Management Service Center, Washington Office (Detached)<sup>1</sup>

Phil R. Krueger, Forester  
Medicine Bow National Forest<sup>2</sup>

<sup>1</sup> Located in Fort Collins, Colorado. Donnelly was formerly Forester with the Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.

<sup>2</sup> Located at Supervisor's Office, Laramie, Wyoming.



## CONTENTS

	Page
INTRODUCTION .....	1
OBJECTIVES.....	1
SAMPLING THEORY.....	1
Steps in Sampling .....	2
Sampling Error and Risk .....	3
Sampling Scheme .....	4
Simple Random Sampling (SRS) and Weights.....	4
Sampling with Probability Proportional to Size (PPS).....	4
Comparing SRS using EPS and SRS using PPS .....	5
APPROACH TO SAMPLING .....	7
Sample Selection .....	7
Sample Size .....	10
Allocating the Sample .....	11
Under- and Over-Sampling.....	11
List of Samples.....	12
SAMPLING FOR PLANNING .....	13
Sample Stand Selection Based on RMRIS Site Maps .....	14
Use of the Stand Selection Procedure .....	17
REFERENCES CITED .....	22
APPENDIX .....	22
Text of File SELECT_STANDS.README .....	22
Text of File SELECT_STANDS.FOR .....	25
Text of File SELECT_LP9.SQL .....	27
Partial Text of File SELECT_LP9.LIS .....	27



# Selecting Stands for the Forest Planning Data Base: Sampling Background and Application

Dennis M. Donnelly and Phil R. Krueger

## INTRODUCTION

Sampling combines science in the form of statistics and mathematics with art in the form of creativity, judgement, and approximation. The science of sampling helps determine the effects resulting from deviations from absolute rigor. The art of sampling helps the user know whether or not these deviations are acceptable. Most applied sampling situations require some compromise between total rigor and the allocation of available resources. This "trade-off" typically balances costs (e.g., funds, time, effort) and sample attributes (e.g., precision, representativeness) within a framework of objectives which establishes boundaries for both science and art. Cochran (1977) discusses these ideas in greater detail in his Preface and Introduction (Chapter 1).

The sampling methods described in this Report are derived from the Forest Service's Resource Inventory Handbook (USDA Forest Service 1990), which explains in its introductory paragraphs the purpose for sampling and related analysis activities:

"Periodic information is required for all land, soil, timber, forage, water, air, fish and wildlife, aesthetics, recreation, wilderness, and energy and mineral resources on all forest and range-lands in the United States for developing the Resources Planning Act (RPA) assessment, program, and subsequent Regional guides and National Forest plans. Resource inventories provide much of the required information . . .

"Coordinated or integrated resource inventories provide efficient, compatible, and valid data and information that describe the resources and their conditions, potential, and trends. Information from the inventories may provide input to the Resources Planning Act (RPA) national assessment, National Forest plans, comprehensive State-wide forest plan assessments, and may be used for project planning where such data are appropriate. Coordinated or integrated resource

inventories promote data sharing among resource managers and decision makers."

For purposes of this Report, the resource considered is forest vegetation, and sampling is focused on estimates of cubic volume of wood which serve as input data to "yield tables" for FORPLAN (Iverson 1986).

## OBJECTIVES

This report reviews principles of sampling theory relevant to sampling forest attributes within the Forest planning context and illustrates how sample stands may be selected. The Sampling Theory section of this report provides guidelines to better understand the approach taken in applying the stand selection procedure. The Approach to Sampling section describes in some detail how the Medicine Bow National Forest applied these sampling principles to select portions of their Forest data base for planning. A central part of the sampling work used in the Medicine Bow National Forest example is a stand selection procedure developed by Dan Greene (Dolores Ranger District, San Juan National Forest). This procedure is currently available as an information and computer routine package, retrievable from the R2 Regional Office in Lakewood, Colorado<sup>3</sup>. This procedure selects stands from a Forest Resource Information System (RIS) data base which will become the Forest planning data base. The appendix provides details about this stand selection procedure.

## SAMPLING THEORY

The related problems of stratification and sampling are often encountered during the analysis of forest data bases for planning. Even though stratification (in the planning sense) is not a focus of this paper,

<sup>3</sup> See the appendix for information required to retrieve these files.

we need to establish its relationship to sampling. For stratification to be effective, the proposed stratification scheme should be meaningful in terms of—1) usefulness in analyzing potential management objectives; 2) availability of data that is describable analytically and can support the stratification analysis; and 3) variation within a stratum which is small relative to the variation existing in the unstratified data.

Stratification of sites may help the analysis when it is intuitively or analytically obvious that categorization of similar sites into groups can reduce overall variation within groups of data. Stratification by cover type and size class seems intuitively obvious. Stratification by density is less obvious but is usually helpful since density is related to wood volume, wildlife habitat and other resource attributes. Stratification by productivity measure may be useful if it can be shown that differences exist between productivity classes. Once strata are defined, the question is how one should sample to reliably depict conditions within strata for a variety of resource attributes potentially useful for planning.

For example, net cubic volume is often considered in planning. Net cubic volume estimates should become less variable when stratified by one or more of the three characteristics mentioned earlier (i.e., cover type, size class, or density). Sampling for one attribute, such as cubic foot volume, does not necessarily preclude analysis (statistical or otherwise) of other attributes available along with cubic volume, such as tree crown structure or canopy closure status of stands. Further, as ecological management becomes the frame of reference within which resource managers conduct analysis and planning, sampling of several attributes simultaneously may become commonplace.

Usually, the goal of sampling is to estimate the mean or total magnitude of a selected measurable attribute over a specific forest area, such as a National Forest. However, in the planning context, it is important to know more than just the overall Forest mean or total value for some selected attribute. The planning process requires that estimates be available for subunits of the Forest (called by various names, such as levels, strata, management areas, ecological land units, etc.). Estimates for these sub-units are equally important for parts of the planning process as are estimates for the whole Forest. Estimates for subunits, such as strata, are used frequently in various types of "yield tables" that are part of the input data requirements for the FORPLAN linear programming system. Consequently, this discussion of sam-

pling theory includes some background about the relationships between overall estimates and subunit estimates based on stratified sampling.

Planning analysis considers the current conditions of land and its vegetation cover, as well as estimating how these resources react to proposed management actions. In this context, the acre is often the fundamental unit of area. The acre would be the logical sampling element if data were kept on an acre basis, but this is almost never done. Instead, data are taken by sampling points within stands. Thus, the stand is the sample element and the sampling unit for purposes of planning analysis. See Mendenhall et al. (1971, p.20), Cochran (1977, Chap.1), or any other comparable sampling text, for more information on terminology, such as "sampling element" or "sampling unit."

For sampling purposes, the total population of stands includes all potential sampling units on the Forest. In past and current practice, some stands, such as those in wilderness areas, are deleted from the total population because they are legally withdrawn from consideration for standard management. The remaining net population of sites with forest cover are grouped into strata that are the basis for data gathering and analysis. However, in the very near future it may become necessary under ecological management approaches to sample stands in all areas of a National Forest. Such widespread sampling might involve collecting data on a suite of stand variables which account for ecological and landscape-scale interactions, regardless of the legal designation of National Forest land.

## Steps in Sampling

The following steps are broad guides for preparing a sampling design based on the generally accepted principles for statistical sampling.

1. Decide how much "error" can be tolerated in the estimates of the attribute of interest. "Error" in this case refers to the statistical boundaries needed for precision. For example, an error specification could be stated as follows: estimate the average net cubic foot volume per acre to within  $\pm 10$  percent of its "true" but unknown value (USDA Forest Service 1990:Sec. 25.3). It may not be feasible to specify statistical "error" so precisely for many of the hundreds of data categories needed for planning, because so

many kinds of information are derived from so many sources with varying degrees of precision. However, volume can usually be measured with such standards because many tree-related measurements are relatively precise.

2. Decide what acceptable "risk" can be tolerated. Specify the probability that the estimate is, or is not, within the designated error boundaries. For example, given the above specification for allowable "error", decide that a 67 percent chance would exist that the estimated mean is within 10 percent of the unknown actual mean (USDA Forest Service 1990:Sec. 25.3). This figure indicates that the estimated mean values from *repeated samples* would fall within the specified error bounds around the true mean in 67 percent of the sampling cases.
3. Estimate the variability of the subject population. The population variability and the selected sampling scheme greatly influence how many samples must be taken.
4. Decide on the sampling scheme. This choice depends on tradeoffs between cost and precision requirements, on the statistical properties of the attribute(s) to be measured, and on the organization of the system within which the sampling is to be done. The "system" to be sampled, for example, might be the lands within a National Forest designated as "suitable" for resource management, as opposed to those lands withdrawn from management, such as wilderness areas.
5. Based on the sampling scheme and on the desired levels of precision and cost, estimate how many samples must be taken.
6. Decide how the data should be collected, i.e., field work, access to a data base, etc.

For additional detail on sampling, see Cochran (1963, 1977). Freese (1962) is also a valuable reference.

### **Sampling Error and Risk**

Considering the almost infinite variety of information that could be sampled across the resources available on a National Forest, it follows that guidelines for such sampling are also highly diverse. This is evident in the Forest Service Resource Inventory Handbook (USDA Forest Service 1990). Because wood products are traditional National Forest outputs, sampling and inventory methods for tree vol-

ume are well developed, and very specific direction is given about precision guidelines for data sampling and analysis. In addition, the Forest Service Handbook 2409.13, "Timber Resource Planning Handbook" documents procedures for inventory and analysis (USDA Forest Service 1992). Sections 12.1, "Precision Requirements", and 43.1, "Inventory Data", of this Handbook suggest that growing stock volume in gross cubic feet should be estimated with a sampling error of  $\pm 10$  percent at the 67 percent level of statistical confidence (i.e., there is a two-in-three chance that the estimate is within  $\pm 10$  percent of the unknown population mean). Section 12.1 also states: "There are no other national requirements for sampling accuracy (or precision), as long as the method of collecting the data is objective so that it is possible to calculate sampling errors should the need arise." But, "Regional Foresters shall supplement these standards relevant to the local decisions." The reader should consult this Handbook and Regional or local supplements for other details that may affect the application. For example, the 10 percent sampling error is listed for all Forest Service Regions except the Southeast (Region 8) and the Northeast (Region 9). In these regions, the allowable standard error is five percent applied to growing stock volume on Forest Land Not Withdrawn.

When dealing with statements about statistical means and other measures of resource attributes, it is often helpful to think in terms of statistical confidence intervals (Dixon and Massey 1969, Sokol and Rohlf 1969). The following summary points also draw heavily on Freese (1967).

1. A statistical confidence interval (CI) based on a normally distributed population is given by the relation:

$$(\text{estimate}) \pm (t)(\text{standard error})$$

where the estimate is typically an estimated mean value, "t" is a value with probability p and degrees of freedom d from the t-distribution, and the standard error is the sample estimate of the standard deviation.

2. In simple sampling cases, the degrees of freedom measure is computed as sample size minus 1, or  $d = n - 1$ . Thus, for large samples, say 40 or more, degrees of freedom is almost the same value as the sample size. The probability p is a measure of the probability that the confi-

dence interval includes the true mean. The typical 95 percent confidence interval expresses the idea that the CI contains the mean 19 times out of 20. A 67 percent CI likewise expresses the idea that the CI includes the mean two out of three times.

3. Sample estimates and their confidence intervals depend on the amount of variation in the population, the size of the population, and the number of samples taken.
4. The effectiveness of CIs to provide information about an estimate is a tradeoff between precision and probability. All else being equal, increasing precision (i.e., reducing the interval around an estimate) implies decreasing the probability that the interval actually contains the estimate.
5. Certain rules of thumb follow from these definitions, and from consulting the references and a table of t-values. For "large" samples, say 40 or more, the t-values involved can be approximated for common CI percent limits:

$$99\% \text{ (estimate)} \pm (2.7)(\text{standard error})$$

$$95\% \text{ (estimate)} \pm (2.0)(\text{standard error})$$

$$90\% \text{ (estimate)} \pm (1.7)(\text{standard error})$$

$$67\% \text{ (estimate)} \pm (1.0)(\text{standard error})$$

Note that the last approximation corresponds to the precision discussed above as required for Forest resource sampling.

## Sampling Scheme

The ideal sampling element to determine forest cover characteristics would be some fixed unit area such as an acre. If, for example, it were possible to measure all trees on randomly selected acres over a large area, with each such acre equally likely to be selected for the sample, then unbiased estimates of quantities such as cubic volume per acre could be easily computed. However, data taken by the whole acre are usually not available. What is available are data collected on plots distributed throughout stands (sites) using one of several variable radius or fixed radius plot overlay location schemes, e.g., see Stage and Alley (1972) and Lund and Thomas (1989).

Per-acre estimates computed from these data represent forest stands ranging in size roughly from several acres to several hundred acres. For the purposes discussed in this paper, per-acre estimates for indi-

vidual stands are available in the RIS data bases of the individual National Forests. These estimates are based on several types of stand inventory procedures.

Because stands vary in area, and because Forest planning uses stratification, we must consider the differences between sample organization and choosing sample elements within a particular sample organization. Two methods of sample organization are simple random sampling (SRS) and stratified random sampling (StRS). A key question when considering sample organization is HOW MANY sample elements must be selected to satisfy the objectives of the survey, including precision and cost requirements.

Once an organization method has been chosen, one must decide HOW to pick WHICH sample elements. One may select sample elements with equal probability of selection (EPS) or with a probability proportional to size (PPS). For example, stands could be picked from the entire population of stands delineated on a National Forest, with the choice of stands determined by an equally probable random pick from all stand numbers. Using the acronyms from the previous paragraph, this is SRS with EPS. In contrast, the Forest could be stratified on the basis of species and size class, and the stands chosen within each stratum with a probability based on the area of each stand, i.e., StRS with PPS. Other combinations are possible.

## Simple random sampling (SRS) and weights

If a sample of stands is randomly chosen from the RIS data base or even if all stands in the RIS data base are used, it is often necessary to weight individual stand estimates by the number of acres in the stand to offset the effect of widely varying stand size. How the weighting is done depends on the nature of the estimate, the particular question to be answered, and the structure of the sampling scheme. For example, if estimates of cubic volume per acre from 100 randomly chosen stands are averaged without weighting, then the volume-per-acre figure from a 5 acre stand counts just as heavily in the average as the volume-per-acre figure from a 500 acre stand. Weighting by stand area is especially useful when the sites involved have highly unequal areas.

## Sampling with probability proportional to size (PPS)

An alternative sampling scheme for weighting relative to stand size is called sampling with probability of selection proportional to size (PPS) (Freese

1962, p.47-50; Cochran 1977, Chap 9A). This approach is the basis for the stand selection procedure used for the planning data base described in this report. It should be noted that a related procedure based on PPS is sampling with probability proportional to prediction (PPP or 3P sampling) (Grosenbaugh 1965; Lund 1975). However, in the application considered in this paper, 3P sampling should not be confused with PPS sampling.

PPS sampling is typically done by setting a limit or interval for the weighting variable against which accumulation of weights is compared (Freese 1962, p. 47-8; Cochran 1977, p. 250-1). When the sum of weights from each potential sample element reaches or exceeds the limit, the sample element whose weight exceeded the limit is chosen to be part of the sample. In the PPS stand selection procedure, the limit typically ranges from 1000 to 3000 acres, and each stand area in acres is the weight that is accumulated. How many samples are picked is determined by the magnitude of the limit. The number of samples needed is addressed later.

Another important point about sampling regards sampling with replacement and without replacement. Most of the ideas discussed so far assume sampling with replacement. Cochran (1977:p.251) notes that variance formulas for PPS sampling with replacement are relatively simple. However, the stand selection procedure is sampling without replacement, because once a stand is picked as part of the PPS routine, it is not eligible to be picked again. But as sample size decreases relative to population size, for example less than 10 percent, the chance of sampling the same stand again is low. Thus, we assume that when sample size is small relative to population size, sampling without replacement is approximate to sampling with replacement.

### Comparing SRS using EPS and SRS using PPS

This example illustrates how Simple Random Sampling (SRS) and sampling with Probability Proportional to Size (PPS) work and compare with each other. Figure 1 and table 1 contain the information used in this example.

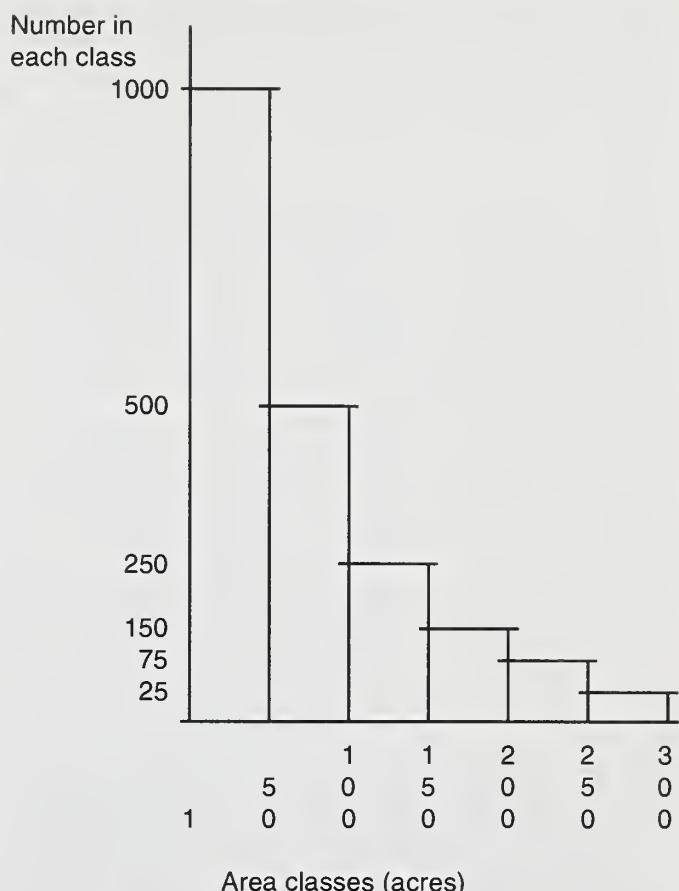


Figure 1.—Histogram showing distribution of stands by their area in acres for the hypothetical example comparing Simple Random Sampling (SRS) and sampling with Probability Proportional to Size (PPS).

size classes, the distribution of stand areas is shown in the histogram in figure 1.

1. Assume a planning stratum has 2000 stands—the smallest stand is 1 acre and the largest stand is 300 acres. Assume that, as in real situations, the stand sizes are distributed negative exponentially as stand size increases. Thus, if we group the hypothetical stand sizes into 50-acre
2. Using figure 1 as a base, we begin forming the information in table 1. Column 1 lists the stand area size classes and column 2 lists the number of stands in each class. This is the same information contained in figure 1.
3. Column 3 is the proportion of the total number of stands within each area size class. Column 4 is a hypothesized average stand area of the stands in each size class. Column 5 lists the total acres in each area size class, and is the product of the numbers in column 2 and the corresponding numbers in column 4.
4. Assume that a desirable sample size is 10 percent of the total number of stands in the stratum (a discussion about sample size follows this section). The expected distribution of sampled stands by area size class is shown in column 6.

**Table 1.—Comparison based on hypothetical data of Simple Random Sampling and sampling with Probability Proportional to Size.**

Column No.→		(01)	(02)	(03)	(04)	(05)	(06)	(07)	(08)	(09)	(10)	(11)
Size class (acres)	Number of stands	Proportion of total	Average stand size (acres)	Total in size class (acres)	Simple random sample: 10% of stands	Total sample area (acres)	Each sample area proportion of total (7)/sum(7)	Compute samples for PPS sum(5)/sum(6)	PPS samples: proportion of sample size (9)/sum(6)	PPS samples rounded (9) rounded	PPS samples rounded (9) rounded	
1–50	1000	0.5000	15	15000	100	1500	0.1186	23.72	0.120	24		
51–100	500	0.2500	65	32500	50	3250	0.2569	51.38	0.255	51		
101–150	250	0.1250	120	30000	25	3000	0.2372	47.43	0.235	47		
151–200	150	0.0750	170	25500	15	2550	0.2016	40.32	0.200	40		
201–300	75	0.0375	225	18000	8	1800	0.1423	28.46	0.140	28		
251–300	25	0.0125	275	5500	2	550	0.0435	8.70	0.045	9		
	2000	1.0000		126,500	200	12,650	1.0000	200.00	1.000	199		

In our hypothetical case, the sample distribution is exactly the same as the population distribution. With actual stands, one single sample drawing may or may not have a size distribution that approximates the size distribution within the whole population. However, if a sample were selected from the population, then returned to the population before the next selection, and this process were repeated 50 or 100 times, the statistical expectation is that when all the samples are considered, the average number of stands in each size class would closely approximate the size class distribution of the whole population.

- Given our 10 percent sample, with distribution of stands as shown in column 6, the stand area sampled from each size class is shown in column 7. Here is a key point for SRS: based on the number of stands in the sample, the smaller size classes are represented to a much greater degree than the larger size classes (column 6 in table 1). The 100 stands in the smallest size class (half of the SRS) have only about 12 percent of the total area in the sample (column 8). This is why weighting by acreage is important when making inferences based on a simple random sample as shown in this example.

How would this work with PPS? Lund (1978) provides a comprehensive example of PPS sampling, il-

lustrating both the science and art. Cochran (1963, 1977:Chap 9A) presents the theoretical background on this topic. At the beginning of this paper we mentioned science and art in sampling. In our example, our objective is to obtain a comprehensive sample that represents the range of conditions in the stratum. Thus, we can make approximations suited to the resources available for our sampling and still stay roughly within the bounds established by sampling theory.

- To determine the cutoff number of acres to be compared with the cumulative sum of acres from stands in the population, Lund (1978) demonstrates an approach that is used in Greene's procedure in slightly modified form. This modification is demonstrated in column 9 of table 1. We need the average area in acres in the population that will be represented by each stand in the sample. This is found by dividing the total acres in the population by the total number of stands in the sample (the sum of column 5 divided by the sum of column 6, i.e.,  $(126,500 / 200 = 632.5)$ ). Thus, 632.5 acres is the cutoff number of acres for this example. We have approximated the Lund method by simply dividing the number of acres in each size class (column 5) by 632.5. This results in the number of stands which must be sampled in each size class (column 9), and these figures are rounded in column 11.

2. The proportion of stands in each size class in the sample is computed in column 10. This is the key point for PPS sampling. The proportions in column 10 closely match the proportions in column 8. In effect, this incorporates the weighting required in SRS into the sample that is sufficient for PPS sampling. Due to rounding, the total number of samples is 199, not 200 as shown in column 11.

In our example, the proportion of area (not number of stands) in each size class is about the same as the proportion of area in the population in each size class. That is the result we are seeking and the approximate result obtained from using the PPS stand selection procedure described in the appendix.

## APPROACH TO SAMPLING

Given the sampling concepts discussed so far, the classic situation focuses on determining the number of samples for a study and how to select them. The theoretical situation is usually clear. The goal is usually to take the minimum number of samples that will allow inferences to be drawn within prescribed statistical error bounds. Alternatively, it is also valid to take a certain predetermined sample size and minimize the error. These objectives often apply to narrowly- and well-defined sampling problems. However, consider the sampling problem within the context of Forest Planning. Forest stand information, especially cubic volume information, is merely one class of information among a much more diverse set of topics that could conceivably enter into planning analyses. Even within the class of forest stand information, many attributes may be needed for planning analysis, such as canopy cover or vertical diversity.

For now, we want a representative sample within each stratum, considering only cubic volume. At least two philosophical approaches are available for picking the sample. The first is to consider each stratum as an independent unit. Here the sample size for each unit is picked independently of any other unit. This is simple random sampling applied to each stratum. The stratum results may then be aggregated to the Forest level. The second approach is to consider each planning stratum as part of a larger integrated Forest Unit for statistical purposes. This is stratified random sampling. Just as for simple random sampling, strata statistics may be aggregated, if necessary, to the Forest level.

One must decide whether to determine sample size at the stratum level, based on stratum attributes, or to determine sample size at the Forest level and then allocate the sample to the various strata. If the sample is determined at the Forest level so that it satisfies the minimum precision requirements for Forest-wide volume estimates, there is no guarantee that the sample, when allocated among the strata, will satisfy the same minimum precision requirement at the stratum level.

When a Forest-wide sample is allocated to several strata on the basis of area, as is done in this stand selection process, it may be assumed that variation within each stratum is proportional to area, i.e., the larger the stratum area, the more variation there is within the stratum. However, while this seems intuitive, there really is no reason to expect this condition. One reason why strata are formed is to create groups of stands that are relatively more homogeneous (having less variation) than would otherwise be the case.

Thus, in the case of strata with large areas, it is possible that the allocated sample may exceed reasonable requirements ("over-sampling"). It may be possible to reduce the allocated number of samples for large strata. Conversely, only a few samples may be allocated to strata with small areas ("under-sampling"). In the latter case, the "rule-of-thumb" is to take a minimum of five samples. Both these cases should be considered in the context of planning requirements. If changes are made to sample sizes because of over- or under-sampling, one should be aware of the consequences of such adjustments, both for estimates of individual strata and for aggregates of stratum estimates to some larger area, such as a District or Forest.

Now let us look at the sampling process for Forest planning in more detail. The usual approach for picking samples within each stratum is based on Stage (1971), as implemented in the stand selection procedure. See the appendix of this report for an example. In order to sample stands in Forest areas suitable and available for management and harvest, strata are defined based on a minimum of species and size class, and a sample size is determined which satisfies Forest-wide precision requirements, assuming the use of stratified random sampling. Specific sample elements within each stratum are picked with probability proportional to size.

## Sample Selection

The task is to select stands from the list of several thousand within each stratum that make up the for-

ested area available for management. Summary data from the Medicine Bow National Forest will help demonstrate the approach. The data presented here are sketchy, illustrating that sometimes it is possible to go ahead with a first approximation even when the ideal amount of information is not available. Based on these data, levels of "error" and probability are chosen so that results approximate conditions used in the stand selection sampling procedure. The purpose of this small-scale exercise is to test the procedure for computing and choosing sample size using limited but realistic data.

PPS sampling is implemented within each stratum defined by cover type and size class. Though this sample is not intended to be a basis for statistical in-

ference about Forest-wide parameters, we assume that the sample satisfies the probability that a sampling error of approximately  $\pm 10$  percent will bracket the true mean of the chosen stand characteristic two out of three times when aggregated over the entire Forest. Stratified sampling with probability of selection proportional to size takes into account stratification by forest cover type and size class, and the variable population distribution of stand area. In figure 2, the distribution of stand sizes in the population is shown by the solid line and is similar to that in the earlier hypothetical example.

The next step in this exercise is to estimate the number of samples needed to achieve the precision requirements of the analysis. For this and for much

## LARAMIE DISTRICT LP-9 COMPARE SP1 SAMPLE SITES TO RIS DATA

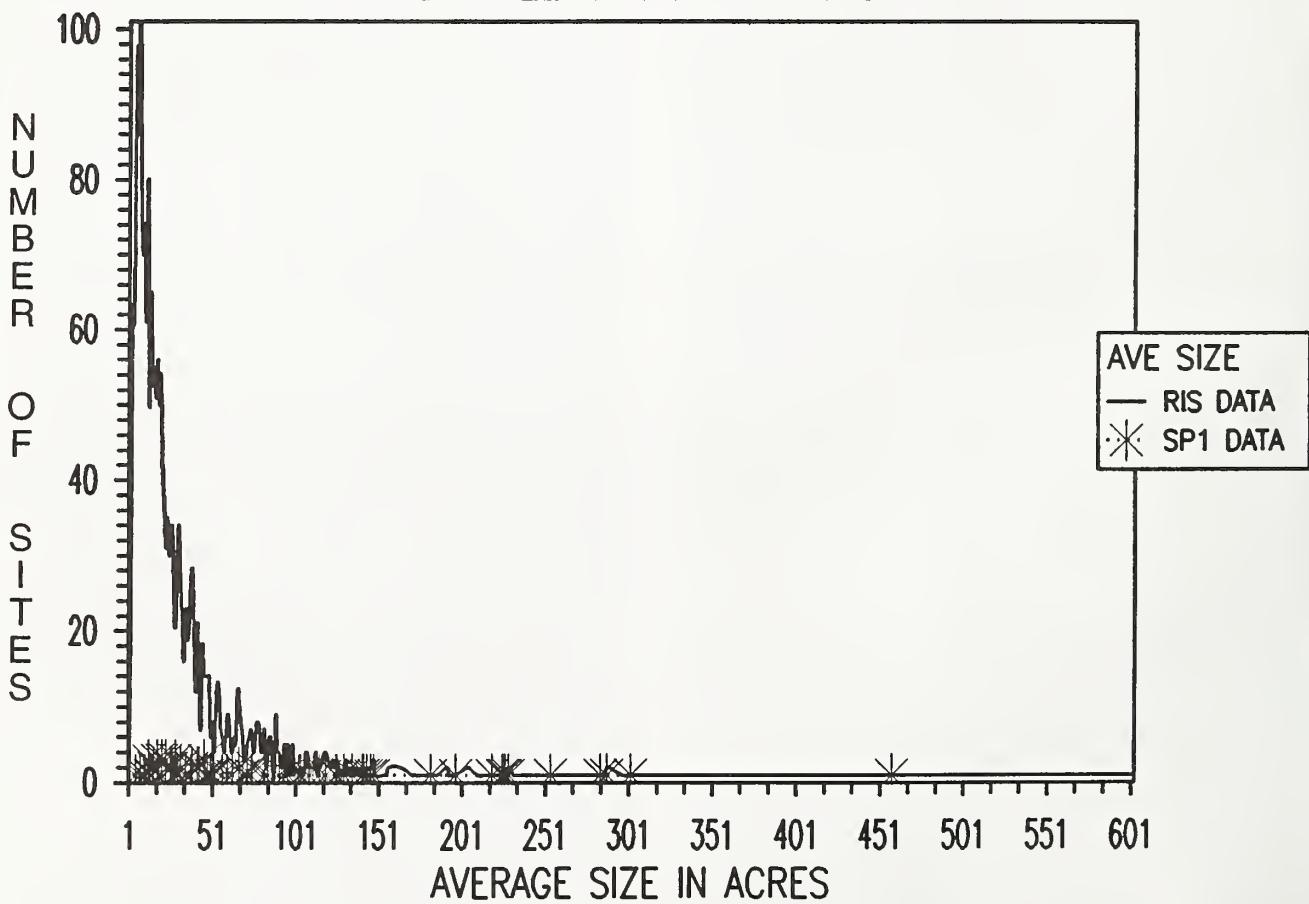


Figure 2.—Frequency of site occurrence based on site size and depending upon selection from the RIS data base or from the sample selection routine based on site area in acres.

of the following discussion, refer to table 2. Because sample number is in part a function of the inherent variability in the data, first compute or estimate variance within each stratum. With the RIS data base, it should be possible to set up a query that—1) extracts cubic foot volume or any other desired continuous variable from all stand records for a stratum; and 2) computes the variance. However, if computing variance for each stratum is not possible, but the range of the values is known, one may roughly approximate variance using the following formula:

$$\text{Variance estimate} = (R/4)^2$$

where the range R equals the largest value minus the smallest value for the desired variable; in this case, cubic volume per acre for stands in the strata.

This estimate applies to populations with a normal or approximately normal distribution because almost all of the sample values over a wide range of sample sizes are bounded within an interval of four to six standard deviations (Dixon and Massey 1969, p.136; Freese 1962, p.25). Using four standard deviations results in the 4 in the denominator of the variance estimate expression above. The estimate of cubic volume per acre computed for each stand often has a normally shaped distribution in contrast with the negative exponential distribution of stand sizes in acres.

Using the data summaries from the Medicine Bow National Forest and the approximation above, column D of table 2 shows initial variance estimates for some of the strata listed. Variance magnitudes for as-

**Table 2.—Spreadsheet format for analysis by strata of variance and sample size.**

STRATA	TOTAL ACRES IN STRATA	INTENSIVE SURVEY ACRES IN STRATA	INITIAL VARIANCE ESTIMATES <sup>1</sup>	REVISED VARIANCE ESTIMATES	TOTAL SAMPLES			
					COL. BxE	1000 STRATUM SAMPLES	MIN OF 5 STRATUM SAMPLES	MAX OF 40 STRATUM SAMPLES
(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)
AA6	2511	1732		20	50220	4	5	5
AA7	4861	2527		20	97220	7	7	7
AA8	35616	15222	128.62	30	1068480	51	51	40
AA9	35729	24803	68.48	70	2501030	51	51	40
DF6	94	39		40	3760	0	0	0
DF8	802	685		60	48120	1	5	5
DF9	5312	4468		80	424960	8	8	8
LP6	13173	2045		20	263460	19	19	19
LP7	57763	20307		20	1155260	82	82	40
LP8	179949	111971	40.51	40	7197960	256	256	40
LP9	197851	151078	79.25	80	15828080	282	282	40
SF6	3490	1263		30	104700	5	5	5
SF7	20152	8542		30	604560	29	29	29
SF8	14371	10060	44.02	45	646695	20	20	20
SF9	130708	91797	60.57	60	7842480	180	186	40
	702382				37836985	1000	1006	338
SAMPLING ERROR (E)		TOTAL N	Example calculation for sampling error = 0.22:					
0.20		1344						
0.22		1111						
0.24		934	(702,382)(37,836,985)					
0.26		796						
0.28		686	((702,382)(0.22)) <sup>2</sup> + 37,836,985					
0.30		598						
0.32		526	= 1111.25, round to 1111					

<sup>1</sup> Units are cunits squared; 1 cunit = 100 cubic feet. Cunits are used to describe variance and desired sampling error because many sample size computations require squaring which otherwise results in large numbers. Using cunits avoids this problem.

pen (AA), lodgepole pine (LP), and spruce-fir (SF) appear similar. To obtain variance estimates for the rest of the strata, we rounded the computed variance estimates to the nearest 5 or 10. Note that the variance for the poletimber category (xx8) is lower than that for the corresponding mature category (xx9). We used this trend to "guesstimate" variances for strata without variances.

Finally, since seedling/sapling (size class xx7) aspen and lodgepole pine stands are relatively uniform, smaller variance estimates were made for these than for either poletimber or mature size classes. The result is shown in column E of table 2 as a hypothetical revised variance estimate. However, variance estimates obviously should be based on real data whenever possible.

### Sample Size

In this section, we discuss several common equations for determining sample size using stratified random sampling. Then we use the last equation to roughly estimate the range for total number of samples required for the Forest Planning task.

The most general equation form considered here is taken from Freese (1962:p.34),

$$n = \frac{(N) \sum_{h=i..L} (N_h s_h^2)}{N^2 D^2 + \sum_{h=i..L} (N_h s_h^2)} \quad (1)$$

where:

$N$  = number of items (e.g., acres) in the population

$s_h$  = the estimated standard deviation of the population attribute (e.g., cunits); thus,  $s_h^2$  is the estimated population variance

$N_h$  = the number of items (e.g., acres) in the population of the "h-th" strata

$h$  = the index denoting a stratum (e.g., LP9 in table 2); ranges in value from 1 to a total of L strata

$D$  = the desired size of the standard error of the mean.

The notation  $\sum_{h=i..L} (N_h s_h^2)$  is the summation over the number of strata between 1 and L, of the product of the number of items in the "h-th" strata multiplied by the estimated variance in the "h-th" strata.

Another form of equation [1] uses Student's t-value to make computation easier. To do this, start with the desired size for standard error of the mean,  $D$ ,

and use the following algebraic trick. Since multiplying any quantity by 1 does not change its value, multiply D by 1, which in this case is set equal to  $t$  divided by  $t$  ( $1 = t/t$ ), giving

$$D = tD/t = E/t$$

where  $t$  is the value from the Student t-distribution, typically used for evaluating "t-tests". The numerator of this expression,  $tD$ , is the same as the "sampling error"  $E$  (the specified margin of error) used by Freese (1962:pp.24,34). Making these changes in equation [1] above gives,

$$n = \frac{(N) \sum_{h=i..L} (N_h s_h^2)}{N^2 (E/t)^2 + \sum_{h=i..L} (N_h s_h^2)} \quad (2)$$

Now we can make some further approximations to illustrate other common sampling cases. Consult a table for "t-values", e.g., Freese (1962, p. 86). Student's t assumes some common values given infinitely large degrees of freedom (df). For example, the column for a probability of 0.05 (1-in-20 chance) provides a t-value of 1.96, approximately 2. In fact, this same t-table column tells us that for degrees of freedom greater than 25, the t-value is approximately 2. Using this value for  $t$  in the formula gives:

$$n = \frac{(N) \sum_{h=i..L} (N_h s_h^2)}{N^2 (E^2 / 4) + \sum_{h=i..L} (N_h s_h^2)} \quad (3)$$

Since degrees of freedom (df) is one less than the number in the sample (n), using equation [3] requires the implicit assumption that n is approximately 25 or greater, and that a 95 percent confidence interval is appropriate for bracketing the estimated unknown mean value.

Another common formula for establishing sample size for simple and stratified sampling (Freese 1962, p. 34) results when equation [2] is transformed based on the assumption that the t-value has a 67 percent probability. In this case, equation [2] becomes:

$$n = \frac{(N) \sum_{h=i..L} (N_h s_h^2)}{N^2 E^2 + \sum_{h=i..L} (N_h s_h^2)} \quad (4)$$

The t-value in the first term of the denominator has disappeared. For n greater than 25 at the 70 percent

level, the t-value is close to 1 (Freese 1962, p. 86), and so is not an explicit factor in the equation. The same approximation applies to t-values at the 67 percent level. Equation [4] applies to the confidence limits specified for inventory precision in the Timber Resource Planning Handbook (USDA Forest Service 1992, Section 12.1), where 67 percent confidence limits are used. From the same t-table, when N is large and probability is 0.33 (1 - 0.67), the t-value is found by interpolation between probabilities of 0.3 and 0.4. So, if 67 percent (or even 70 percent) confidence limits around the estimated mean are acceptable, equation [4] needs only the acceptable "error" E specified, along with N and the variances, in order to compute sample size n. This is the formula for sampling within strata using probability proportional to size (Lund 1978, p. 6).

Equation [4] was used to calculate the two short columns at the bottom of table 2. Here E is the desired allowable sample error in ccf per acre (1 ccf = 100 cf). For example, 0.2 ccf is 20 cubic feet. The calculated columns show that a sample error (E) ranging from 0.32 to 0.22 ccf results in a total sample size (N) ranging from 526 to 1344 stands. Table 2 summarizes the variances, areas in acres, and sums for use in equation [4]. The range of sample sizes computed is approximately the range used in the sample selection process based on the RIS data base (i.e., 600 to 1000). The sample selection process used in this hypothetical example provides a reasonable interval for sample error and acceptable precision for estimates.

The estimated volume of many sawtimber stands on the Medicine Bow National Forest ranges from 1500–3500 cubic feet per acre.<sup>4</sup> When the upper limit of E is  $\pm 10$  percent of any value in this volume range, the table shows that 0.22 to 0.24 cunits per acre are within the specified limits. Thus, when E in equation [4] is between 0.22 and 0.24, the calculated columns at the bottom of table 2 suggest that about 1000 stands is an adequate sample size.

### Allocating the Sample

We want to distribute the 1000 sites among the various strata in table 2 according to their proportional areas. For example, using the area figures in

table 2, column B,  $(35,729 / 702,382)1000 = 51$  sites which should be sampled within the AA9 stratum (column G). In table 2, column H shows the number of samples within each stratum, where each stratum must have a minimum of five samples. Five samples is the "rule-of-thumb" minimum number deemed appropriate for sampling.<sup>5</sup> Given typical sampling procedures, a minimum of five samples is necessary only for the smallest strata. Note also that the total at the bottom of column H is 1006. This total is greater than 1000 because strata which were calculated to have less than five samples (column G) have been increased to the minimum number, five. The next section looks more closely at the options for adjusting sample sizes.

### Under- and Over-Sampling

Three strata are allocated samples in excess of 100, as shown in column H of table 2, and of these, two strata have samples in excess of 200. This results from the sample apportionment by area. Since the objectives of the analysis are first to formulate strata yield tables, and second to achieve precision goals over the entire Forest, it may not be necessary to have so many samples within these large strata. Moreover, any stratum with more than about 40 samples, as allocated by area, may be represented by a maximum of about 40 samples.

Another reason not to over-sample within a stratum is that all the stands sampled within a stratum will be processed in many ways for the various analysis tasks that are part of planning. For example, if a stratum has 200 stands, all of these must be simulated by the Forest Vegetation Simulator system. The time required for 40 FVS simulation runs is much shorter than for 200 runs, and the results from 40 stands are not much less accurate than for 200 stands. For the following empirical procedure, we assume that, even though estimates of sample variance change as sample size within a stratum is reduced from some large number to roughly 40 samples, the estimates do not change enough to significantly affect sample specifications. We assume that the effects on variance estimation are not significant unless the sample size is well below 40, and therefore it is acceptable to reduce large sample sizes based on area within a stratum.

<sup>4</sup> Personal communication from Phil R. Krueger, Medicine Bow National Forest, 9/22/93. Analysis titled "Weighted cubic feet volume data for LP and SF sites on the Med Bow NF . . . based on trees 5.0" Dbh and larger to a 4.0" top."

<sup>5</sup> Personal communication. Ralph Johnson. 3/28/94. On file, Rocky Mountain Station, Ft. Collins, CO.

Look at the effect of sampling 40 sites within the largest strata. Consider the LP9 stratum in table 2 which has 282 samples. If we reduce the sample size within this stratum to 40, will this much smaller sample still accurately portray the volume characteristic of this stratum with acceptable precision given its population variance? To investigate, use a formula for sample size given simple random sampling, since that is the method used within each stratum. Assuming large N relative to sample size n, this formula (Freese 1962, p.26) is

$$n = \frac{t^2 S_x^2}{E^2} \quad (5)$$

Given our assumption of a 67 percent confidence interval, the t value is approximately equal to one, even for sample sizes down to about 25. From table 2, the estimated variance for this stratum is  $s_x^2 = 80$ . Let sample size n equal 40. To find out what sampling error would result for these values, solve equation (4) for E, giving:

$$E = \sqrt{\frac{t^2 S_x^2}{n}} \quad (6)$$

Since  $t=1$ ,  $n=40$ , and  $s_x^2 = 80$ , the expression becomes  $E = \sqrt{80/40} = 1.414$  ccf = 141 cf. Note that for the more typical 95 percent or 99 percent confidence levels, the t-value is greater and so, correspondingly, is E. In any case, the increase in sampling error from  $\sqrt{80/247} = 0.569$  to 1.414 ccf (approximately 57 cf to 141 cf) incurred by reducing sample size from 247 to 40 does not seem to be excessive in terms of the precision required for large scale Forest planning. Also, the 141 cf error for 40 stands is still less than the 10 percent error specification at the 67 percent confidence level called for in the Resource Inventory Handbook (USDA Forest Service 1990). Column I of table 2 shows the adjusted number of samples to come from each stratum.

### List of Samples

Once you have an estimate of the number of samples needed in each stratum (table 2, column H or column I), the next step is to obtain a list of the sites to sample within each stratum. The first requirement is a list of all eligible sites within each stratum

and their acreages. Be aware that the Forest Service RIS data base in Region Two is District-oriented. In addition to any other proportional allocation for strata, samples must also be allocated to the Districts proportional to eligible land area within each District. This additional allocation is not a part of the conceptual example in this discussion but is a part of the detailed example given later in this report.

The Greene procedure is based on an approach by Lund (1978, pp 7-9), which in turn is derived from Stage (1971). To use this procedure, first compute a sampling interval in acres. Then access the list of all stands in the stratum. Each successive stand in the list has its acreage added to a cumulative total until the cumulative total exceeds the previously computed sampling interval number. Choose the stand for the sample whose acreage, when added to the current total, exceeds the sampling interval. Then reset the current sampling interval total to zero and repeat the process until the required number of stands has been chosen.

A couple of examples will help show how the PPS interval sampling routine works. Table 2 is the source of these figures. To sample from stratum SF7 divide 8542 (survey acres in strata) by 29 (stratum samples) to get 294.55, rounded to 295. Whenever the cumulative sum of the acreages in the list of SF7 sites exceeds 295, pick the last stand added to be the sample. To sample from stratum LP9, divide 151,078 by 40 (the reduced number of strata) to get 3776.95, rounded to 3777. Sum the stand acres that have intensive surveys until the cumulative total exceeds 3777 and pick the last site added to the total to be the sample.

In order to compute the sampling interval for a stratum, one approach is to divide the number of acres in the stratum by the number of samples needed from that stratum. However, as noted in table 2, the number of acres in each stratum that are covered by intensive surveys (those with the required data) are usually less than the total number of acres in the stratum. Since the sample can come only from stands with intensive surveys unless additional (and expensive) field sampling is conducted, divide the required sample number into the intensive sample acreage figure.

Using the available intensive surveys assumes that these stands are similar to other stands in the stratum which do not have intensive surveys. This assumption may be tested if permanent inventory plots are available on the Forest. Dixon and Massey (1969, Chap. 8), Sokal and Rohlf (1969, Sections 9.4 and 13.3), and other comparable texts describe the statistical

procedures for testing whether stand attributes from two samples, such as means of cubic volume, can be assumed to come from one population

To perform this test, select a stand attribute, such as gross cubic volume per acre. Identify and select the permanent inventory plots which occur within the stratum of interest. Do the same for the stand examination plots. These two samples are assumed to come from the same population, i.e., all stands within the stratum. Since permanent inventory plots are thought to portray the full range of conditions in the stratum, if the inventory plots can be shown to be statistically similar to the stand examination plots, then stand examination plots adequately represent the stratum even though they are not taken from all possible stands in the stratum. Pick the stand attribute most likely to be used in the planning analysis, such as cubic feet per acre, for the test.

When you have completed the tests and obtained a list of stands, access the RIS data base, extract the needed data from each sample stand record, and create a planning stand data base with this information. This data base becomes the foundation for many follow-up analyses relative to planning questions.

## SAMPLING FOR PLANNING<sup>6</sup>

This section illustrates an actual application, including necessary approximations, given the theoretical underpinnings of PPS sampling. These are the trade-offs between theoretical detail and cost within the planning process, the goals of which are necessarily broad due to the diversity of resources and issues for a National Forest.

For the Medicine Bow National Forest, a set of sample stands may be selected that represents forested, non-wilderness National Forest System lands within the Medicine Bow National Forest. The representativeness of this sample selection depends on several factors, including:

- 1) How much stand examination or permanent plot inventory data exist for the Forest.

- 2) How widely distributed the plots are (stand examination or permanent inventory) across Districts.
- 3) The purpose of the stand examination or inventory process; stand selection criteria for prior examinations or inventories may influence probabilities of selection toward certain types of stands.
- 4) The history of stand examination data collection for the Forest.
- 5) The validity of your area map by forest type and tree size classes, and the accuracy of the data stored in your database.

Because Forest planning covers large areas (several hundred thousand acres to 1+ million acres in size), the selected sample stands must be well distributed to avoid geographic bias, and should have a probability of selection proportional to stand area, to avoid bias due to stand area. In addition, samples are selected within forest cover types and tree size classes. With enough samples, this distribution provides more flexibility to further stratify, as required by situations and issues, after the initial stand selection process is complete. With enough samples, it is easier to maintain the minimum of 5 samples selected within any single stratum.

Before creating a Planning data base, check with the Regional Office to determine if a valid data set of permanent plots or stand samples has already been compiled based on the latest forest inventory. Past Forest-wide (Stage I in inventory jargon) inventories have generally collected around 300 sample plots to represent a Forest. However, experience has shown that 600 to 1000 samples are more desirable for Forest Planning needs. The increased number allows for allocation to a set of strata that adequately reflect ecological diversity at the strategic level, and also the issues and situations affecting management.

One effective way to sample stands proportional to area, and to meet distribution concerns, is to place a grid over the forest. With a grid, larger stands have a greater chance of selection and the nature of the grid insures distribution. However, you will find that some selected stands have not been sampled, so you must either collect data for those stands, or choose a denser grid.

An alternative method for sampling stands proportional to area is the method we used, a computer-based routine designed to sample stands proportional to area. Distribution across Districts may be

<sup>6</sup>Until now, we have used the term "stand" to refer to more or less homogeneous groups of forest trees. However, Region Two has adopted the term "site", as in forest site or grassland site, to refer more precisely to land areas with a wide variety of potential cover types. This meaning for "site" is also important in the context of the new Integrated Resource Inventory system. Since this part of the report describes application of stand (i.e., site) selection in Region Two, the two terms are used interchangeably; stand = forest site.

assured by selecting separately for each District. Geographic distribution of stands within a District is most useful when stands are ordered within locations. The number of samples on a District within a stratum must be proportional to the area of the stratum on the District. One way to check site distribution is to generate a map showing the location of the selected samples.

Across all Districts, set a sampling interval which will select approximately 1000 total stands for the Forest. The ideas behind selecting a sampling interval are discussed earlier in this Report. The number

of stands allocated to a District should be proportional to the area of a stratum within a District within the Forest. You may not get exactly 1000 samples. If you get too many, you can always eliminate some on an interval basis. Although there is no absolutely "correct" number of samples for the purposes described here, generally 600 to 1000 samples should be adequate, depending on how many strata you are dealing with. As the number of categories which define strata increase, the combined number of stratum classes increases exponentially.

### Sample Stand Selection Based on RMRIS<sup>7</sup> Site Maps

The following procedure is the approach used on the Medicine Bow National Forest.

1. Determine strata, initially by cover type and tree size.
2. Query each District's database to find total acres, number of sites, and average site size by stratum for National Forest System acres that are forested non-wilderness. An example query (assuming use of Oracle-based query language) for each District is:

```
SET ECHO ON
SPOOL XXXXXXXXXX (where X...X is a filename)
SELECT COVER_TYPE, TREE_SIZE, SUM(AREA), COUNT(SITE), AVG(AREA)
FROM R2RIS_SITE WHERE PROC_FOREST = '06' AND
OWNER = 'NFS' AND
MANAGEMENT_AREA NOT LIKE '%8%' AND
COVER_TYPE LIKE 'T%' GROUP BY COVER_TYPE, TREE_SIZE
ORDER BY COVER_TYPE, TREE_SIZE;
SPOOL OFF
```

This query to the RIS data base for the Laramie Ranger District results in the data displayed in table 3. This query process is repeated for each Ranger District.

3. Next, set up a district and forest summary including all the information obtained above for acres and sites for each stratum and for each Ranger District. This will help determine the total acres on the forest and the acres in each District-stratum combination and the percentage of area for each cover type and tree size class

combination. One format for this summary appears in table 4. Notice the minor acreages in cover types TCW, TGO, and TRJ. Small amounts of TDF and TLI exist—one percent and two percent of the Forest total respectively.

4. Now we need to find out how much of the total Forest area is covered by intensive stand exams. Since each District probably does not have intensive data for all sites, query each District as to total acres, number of sites, and average site size by strata for National Forest System forested, non-wilderness acres that have intensive

<sup>7</sup> The Rocky Mountain Resource Inventory System (RMRIS or simply RIS) is a relational database accessed by using Oracle Sequential Query Language installed on the Forest Service's computer systems. These examples are couched in software syntax specific to these systems. However, the general idea should be apparent to anyone. People familiar with relational databases in general could translate queries like these to their own systems.

surveys. The following example query also applies to the Laramie District:

```

SET ECHO ON
SPOOL XXXXXXXXXXXX
SELECT A.COVER_TYPE,A.TREE_SIZE,B.SURVEY_METHOD,
SUM(A.AREA),COUNT(A.SITE),AVG(A.AREA)
FROM R2RIS_SITE A,RMSTAND_HEADER_DATA B WHERE
A.PROC_FOREST = '06' AND
A.LOCATION = B.LOCATION
AND A.SITE = B.SITE AND
A.OWNER = 'NFS' AND
MANAGEMENT_AREA NOT LIKE '%8%' AND
COVER_TYPE LIKE 'T%' AND B.SURVEY_METHOD = 'I'
GROUP BY A.COVER_TYPE,A.TREE_SIZE,B.SURVEY_METHOD
ORDER BY A.COVER_TYPE,A.TREE_SIZE,B.SURVEY_METHOD;
SPOOL OFF

```

The output from this query for the Laramie Ranger District is displayed in table 5.

5. When all the District data bases are queried, the results can be displayed in a District and Forest summary with information for acreages and sites within each stratum for which there are intensive stand data. From this, you can determine the total acres on the Forest by District-stratum combinations and the percentages for each cover type and tree size which have intensive surveys. Table 6 displays this information.
6. Determine the sampling ratio Forest-wide to obtain a total of 1000 samples over all strata. For each stratum, allocate the number of samples to Districts based on the stratum acres within each District. Here is an example:

Using table 4, note that total acres on the Forest is 814,931 acres, and that total TLP acres on the Forest (from the first summary of all acres) is 448,736 acres. Therefore, cover type TLP =  $448,736/814,931$  or 55 percent of total NFS, forested, non-wilderness acres. Thus, 55 percent of 1000 samples equals 550 samples allocated to the entire TLP cover type. The subset of total TLP acres within the Forest LP9 designation is 197,851 acres, which comprises  $197,851/448,736$ , or 44 percent of the total TLP cover type on the Forest. So we must allocate 242 samples (44 percent of the 550 TLP sample sites) to the total Forest LP9 acres. Now look at the Laramie District which has 73,446 LP9 acres. This is  $73,446/197,851$  or 37 percent of the total Forest LP9 acres.

So we must allocate 37 percent of the 242 samples (89 samples) to the Laramie District LP9 Strata.

A shorter way of arriving at the same answer, and also a check, is to compute number of samples for the Laramie district as a proportion of Laramie District LP9 acres to all Forest acres:

$$(73,446/814,931) \times 1000 = 90 \text{ samples}$$

**Table 3.—Number of sites, their cumulative area, and the average site area by cover type for the Laramie Ranger District, Medicine Bow National Forest.**

COV	T	SUM(AREA)	COUNT(SITE)	AVG(AREA)
TAA	6	1046	61	17.147541
TAA	7	1367	68	20.1029412
TAA	8	6153	351	17.5299145
TAA	9	3199	164	19.5060976
TDF	6	83	4	20.75
TDF	8	458	15	30.5333333
TDF	9	2582	68	37.9705882
TLI	6	37	2	18.5
TLI	7	142	3	47.3333333
TLI	8	296	20	14.8
TLI	9	1267	51	24.8431373
TLP	6	5588	387	14.4392765
TLP	7	27285	1198	22.7754591
TLP	8	73025	1903	38.3736206
TLP	9	73466	2215	33.1674944
TPP	6	2229	40	55.725
TPP	7	111	4	27.75
TPP	8	407	12	33.9166667
TPP	9	7654	145	52.7862069
TSF	6	847	32	26.46875
TSF	7	1953	90	21.7
TSF	8	3371	125	26.968
TSF	9	32894	959	34.3003128

Table 4.—Distribution of all suitable acres by cover type and tree size class for NFS, forested, nonwilderness lands (as of 02/25-26/92)

COV. TYPE SIZE	BRUSH-CREEK		LAR.PK. AREA		HAYDEN		LARAMIE		TOTALS	
	SITES	ACRES	SITES	ACRES	SITES	ACRES	SITES	ACRES	SITES	ACRES
TAA	6		2	46	33	1,419	61	1,046	96	2,511
	7	40	977	5	93	2,424	68	1,367	166	4,861
	8	132	4,000	141	1,643	23,820	351	6,153	1,133	35,616
	9	124	3,431	66	944	28,155	164	3,199	824	35,729
sum	296	8,408	214	2,726	1,065	55,818	644	11,765	2,219	78,717
TCW	8		11	89	4	56			15	145
	9	1	10	4	30	254			9	294
	sum	1	10	15	119	310			24	439
TDF	6	1	11				4	83	5	94
	8	4	50	16	272	2	22	15	458	37
	9	13	474	34	1,012	33	1,244	68	2,582	148
	sum	18	535	50	1,284	35	1,266	87	3,123	190
TGO	5				1	46			1	46
TLI	6		2	12			2	37	4	49
	7		2	174			3	142	5	316
	8		199	7,484			20	296	219	7,780
	9		114	3,843	2	16	51	1,267	167	5,126
	sum		317	11,513	2	16	76	1,742	395	13,271
TLP	6	277	3,807	12	220	421	3,558	387	5,588	1,097
	7	862	18,713	5	207	860	11,558	1,198	27,285	2,925
	8	604	25,218	1,642	49,359	608	32,347	1,903	73,025	4,757
	9	1,065	36,606	276	10,154	1,668	77,625	2,215	73,466	5,224
	sum	2,808	84,344	1,935	59,940	3,557	125,088	5,703	179,364	14,003
TPP	6		73	2,759			40	2,229	113	4,988
	7		64	1,157			4	111	68	1,268
	8		82	1,756	1	18	12	407	95	2,181
	9	2	47	2,548	82,585		145	7,654	2,695	90,286
	sum	2	47	2,767	88,257	1	18	201	10,401	2,971
TRJ	5		1	70	1	70				
TSF	6	76	1,629		131	1,014	32	847	239	3,490
	7	661	14,889		192	3,310	90	1,953	943	20,152
	8	246	7,707	27	479	72	2,814	125	3,371	470
	9	1,671	53,810	38	956	738	43,048	959	32,894	3,406
	sum	2,654	78,035	65	1,435	1,133	50,186	1,206	39,065	5,058
DIST./FOREST SUMS										
SITES		5,779		5,364		5,802		7,917		24,862
ACRES			171,379		165,344		232,748		245,460	814,931

The difference between the two results is due to rounding.

Now decide how to allocate the 89 samples among the acres on the Laramie District for which there are intensive stand examinations. Use table 6. According to the summary of intensive surveys, the Laramie District includes 58,556 acres which are covered by intensive stand exams (1,695 sites averaging 34 acres

in size) for the LP9 stratum. We need 89 samples from this stratum. Therefore, the sample selection interval is  $58,556/89$ , or 658 acres. If you base the selection interval on 660 acres you will get very close to the 89 samples you need. Later, if you want to reduce this number to the maximum 40 samples, the sampling error can be checked to be sure it does not increase appreciably as was discussed in the Approach to Sampling section.

**Table 5.—Number of sites with Intensive Surveys, their cumulative areas, and the average site areas by cover type for the Laramie Ranger District, Medicine Bow National Forest.**

COV	T	S	SUM(A.AREA)	COUNT(A.SITE)	AVG(A.AREA)
TAA	6	I	736	44	16.73
TAA	7	I	1227	57	21.53
TAA	8	I	2599	122	21.30
TAA	9	I	2157	88	24.51
TDF	6	I	39	3	13.00
TDF	8	I	390	12	32.50
TDF	9	I	2143	56	38.27
TLI	6	I	37	2	18.50
TLI	7	I	25	1	25.00
TLI	8	I	112	10	11.20
TLI	9	I	1125	41	27.44
TLP	6	I	1427	43	33.19
TLP	7	I	12553	459	27.35
TLP	8	I	62067	1467	42.31
TLP	9	I	58665	1698	34.55
TPP	6	I	1872	36	52.00
TPP	7	I	111	4	27.75
TPP	8	I	247	6	41.17
TPP	9	I	3275	72	45.49
TSF	6	I	583	13	44.85
TSF	7	I	1187	53	22.40
TSF	8	I	2608	94	27.74
TSF	9	I	26,886	701	38.35

23 records selected.

7. Next, run Greene's Stand Select procedure for each District stratum (e.g., Laramie District LP9). More detail about this procedure is found in the next section and in the appendix. When you have selected the correct number of samples, load the results into the ORACLE SITELIST TABLE using SQL\*Loader. (Remember that "site" in Regions Two, Three, and Four is synonymous with the more widely used term "stand".) Use the sitelists to extract raw tree data records from each site for analysis using RMSTAND.
8. Once the chosen samples are adequately distributed and proportional to area within strata and between Districts, run the site information through the computer program RMSTAND and verify the results. Then combine all samples within like strata from the different Districts into strata for the entire Forest.

## Use of the Stand Selection Procedure

This section describes how we used the stand selection program with the RMRIS data base to select

samples of NFS, forested, non-wilderness sites on the Medicine Bow National Forest.

See the appendix for a complete listing of the routines in the stand selection procedure. Among the files contained in the RIS STAND\_SELECT Program Module are SELECT\_STANDS.PR. and SELECT\_STANDS.FOR. SELECT\_STANDS.PR is the executable program code which runs the selection program. SELECT\_STANDS.FOR is the Fortran source code which is translated into the SELECT\_STANDS.PR file. File these in the DG Folder from which you will be working and accessing RMRIS. If you want to change the program, you will have to modify the source code first, then re-compile it using the F77 compiler and F77LINK commands. HOWEVER, EXCEPT UNDER UNUSUAL CIRCUMSTANCES, THERE SHOULD BE NO NEED TO MODIFY THIS PROGRAM. In any case, if you print the SELECT\_STANDS.FOR file, you can read the documentation in the program itself which describes how the stands are selected. Additional information is available in the "dump" file from the Region Two Regional Office (See the appendix for DG retrieval information).

To work with the Oracle part of RIS, first check with the Forest or District Systems Manager to make sure you have access to the ORACLE utility on your DG system, including SQLPLUS and SQLLOAD, and that you have been granted USER PRIVILEGES for ORACLE TABLES. The ability to access databases and read or extract data does not necessarily give one the ability to create sitelists. This must be done with EDIT privileges. If for some reason EDIT privilege cannot be granted, the System Manager will have to complete step 7 described below to load the selected sample sites to a site list. The System Manager will need to check whether or not RMRIS\_LOAD\_SITELIST.CTL has been installed before you can complete step 7. Finally, we also recommend that you read at least the first chapter in the SQL\*Loader User's Guide (about nine pages) to become familiar with the program.

The following steps describe the use of the stand select program to obtain a Planning sample data base for the Medicine Bow National Forest.

1. List PROC\_FOREST, LOCATION, SITE, and AREA for each stratum into a file using ORACLE select. The template SQL query, SELECT\_LP9.SQL, shown in the appendix, is customized as follows:

```

SPOOL LP9
SET PAGESIZE 0
SET LINESIZE 40
SELECT B.PROC_FOREST,B.LOCATION,B.SITE,A.AREA FROM
R2RIS_SITE A,
RMSTAND_HEADER_DATA B WHERE
A.PROC_FOREST = B.PROC_FOREST
AND A.LOCATION = B.LOCATION
AND A.SITE = B.SITE
AND SURVEY_METHOD = 'I'
AND COVER_TYPE = 'TLP' AND TREE_SIZE = '9'
AND OWNER = 'NFS' AND
MANAGEMENT_AREA NOT LIKE '%8%'
ORDER BY B.PROC_FOREST,B.LOCATION,B.SITE;
SPOOL OFF

```

A small part of the sample output file LP9.LIS is shown below. The column identification numbers are added here for illustration, but are not normally part of the file.

1	2	3	4	Column	Name
06	200301	0001	5	1	Forest #
06	200301	0004	26	2	Compartment #
06	200301	0005	16	3	Site #
06	200301	0006	8	4	Site size (acres)
06	200301	0009	3		
06	200301	0016	5		
06	200301	0021	12		
06	200302	0002	14		
06	200302	0003	66		

Note that this is a join-select query and selects sites that have a SURVEY\_METHOD equal to Intensive. Also note that MANAGEMENT\_AREA NOT LIKE '%8%' is one means of not selecting wilderness sites. There are other ways of doing this, e.g., SPECIAL\_UNIT != '4'. You may want to try the various ways to delete the wilderness acres to make sure you have the best possible selection.

2. Rename the file with a five character name indicating District and strata. The program takes filenames of up to five characters. For example LALP9 means Laramie District (LA), lodgepole (LP), sawtimber(9), or alternatively, 05LP9 if the District number (05) is used rather than an abbreviation.
3. Execute the DG AOS Screen Editor (SED) to work with this file. Enter SED LALP9, then delete the last three useless lines which the program appends to the file automatically.

4. Run the STAND\_SELECT program and specify LALP9, or whatever you called it, as the input file. THE PROGRAM RUNS BY SIMPLY TYPING THE FILENAME (SELECT\_STANDS.PR) ON THE COMMAND LINE.
5. When prompted by the program, specify a desired AREA interval for the selection proportional to stand area. Enter 660 to select one stand sample for every 660 acres. Output will be sent to a file called STRATA.DAT. Do not change this filename because the SQLLOAD.CTL macro will search for the STRATA.DAT file to load into the Oracle Sitelist table.

The STRATA.DAT file output should look like the following segment:

```

.
.
.
06,201703,0004,LALP6
06,210203,0002,LALP6
06,210504,0006,LALP6
06,210504,0007,LALP6
06,210505,0004,LALP6
06,210804,0003,LALP6
.
.
.

```

6. Load the resulting output into the R2RIS\_SITE\_LISTS table of RIS ORACLE via the RMRIS\_LOAD\_SITELIST.CTL macro using the command:

```

SQLLOAD / newline
control = RMRIS_LOAD_SITELIST.ctl

```

Table 6.—Distribution of acres with Intensive Survey inventory by cover type and tree class size for NFS, forested, nonwilderness lands (as of 02/25-26/93).

COV. TYPE SIZE	BRUSH-CREEK		LARAMIE PEAK		HAYDEN		LARAMIE		TOTALS	
	SITES	ACRES	SITES	ACRES	SITES	ACRES	SITES	ACRES	SITES	ACRES
TAA	6		2	46	22	950	44	736	68	1,732
	7	17	628	2	46	13	626	57	1,227	89
	8	54	2,471	31	385	176	9,767	122	2,599	383
	9	56	2,611	21	333	272	19,702	88	2,157	437
sum		127	5,710	56	810	483	31,045	311	6,719	977
TCW	9	1	10		1	83				
	sum	1	10		1	83			2	93
TDF	6						3	39	3	39
	8	1	22	13	251	2	22	12	390	28
	9	8	295	27	794	31	1,236	56	2,143	122
	sum	9	317	40	1,045	33	1,258	71	2,572	153
TGO	5				1	46			1	46
TLI	6		2	12			2	37	4	49
	7		2	174			1	25	3	199
	8		7	135			10	112	17	247
	9		7	166	2	16	41	1,125	50	1,307
sum			18	487	2	16	54	1,299	74	1,802
TLP	6	11	150	11	211	1	257	43	1,427	66
	7	222	6,526	3	135	39	1,298	458	12,348	722
	8	373	19,000	122	3,409	448	27,495	1,467	62,067	2,410
	9	598	27,283	120	2,815	1,067	62,424	1,695	58,556	3,480
sum		1,204	52,959	256	6,570	1,555	91,474	3,663	134,398	6,678
TPP	6		6	112			36	1,872	42	1,984
	7		6	82			4	111	10	193
	8		24	483			6	247	30	730
	9	2	47	306	8,837		72	3,275	380	12,159
sum		2	47	342	9,514		118	5,505	462	15,066
TRJ	5		1	70					1	70
TSF	6	21	493		4	187	13	583	38	1,263
	7	215	5,791		41	1,571	51	1,180	307	8,542
	8	142	4,970	4	38	52	2,444	93	2,608	291
	9	755	33,957	25	461	459	30,493	701	26,886	1,940
sum		1,133	45,211	29	499	556	34,695	858	31,257	2,576
DIST./FOREST SUMS										
SITES	2,476		742		2,631		5,075		10,924	
ACRES		104,254		18,995		158,617		181,750		463,616
% NFS FOR. INT.		61%		11%		68%		74%		56%

SQL\*Loader: Version 1.0.27 - Production on Wed Mar 3 07:31:46 1993

Copyright (c) Oracle Corporation 1979, 1989. All rights reserved.

Control File: RMRIS\_LOAD\_SITELIST.CTL

Data File: STRATA.dat

Read Mode: System Record

Bad File: RMRIS\_LOAD\_SITELIST.bad

Discard File: none specified

Number to load: ALL

Number to skip: 0

Errors allowed: 50

Bind array: 64 rows, maximum of 65336 bytes

Record Length: 968 (Buffer size allocated per logical record)

Continuation: none specified

Table R2RIS\_SITE\_LISTS, loaded from every logical record.

Insert option in effect for this table: APPEND

Column Name	Position	Len	Term	Encl	Datatype
PROC_FOREST	FIRST	*	,	,	CHARACTER
LOCATION	NEXT	*	,	,	CHARACTER
SITE	NEXT	*	,	,	CHARACTER
SITE_LIST_NAME	NEXT	*	,	,	CHARACTER

Table R2RIS\_SITE\_LISTS:

6 Rows successfully loaded.

0 Rows not loaded due to data errors.

0 Rows not loaded because all WHEN clauses were failed.

0 Rows not loaded because all fields were null.

Space allocated for bind array: 62464 bytes (64 rows)

Space otherwise allocated: 63572 bytes

Total logical records skipped: 0

Total logical records read: 6

Total logical records rejected: 0

Total logical records discarded: 0

Run began on Wed Mar 3 07:31:44 1993

Run ended on Wed Mar 3 07:32:18 1993

Elapsed time was: 00:00:33.69

CPU time was: 00:00:03.26 (May not include Oracle CPU time)

Figure 3.—Sample output from the routine which loads sampled site data into a Forest planning data base.

If the macro has not been installed, create the file RMRIS\_LOAD\_SITELIST.CTL containing the following information:

```
LOAD
INFILE STRATA
INTO TABLE R2RIS_SITE_LISTS
APPEND
FIELDS TERMINATED BY ''
(
PROC_FOREST,
LOCATION,
SITE,
SITE_LIST_NAME
)
```

This macro should reside in the AOS folder from which you are working for easiest access and ease in running the program. The output files created by SQL\*Loader will be RMRIS\_LOAD\_SITELIST.LOG and RMRIS\_LOAD\_SITELIST.BAD. If the run was successful and all sites were loaded, the \*.BAD file should be empty and the \*.LOG file should appear similar to figure 3.

The sitelist should now be loaded into the Oracle SITELIST Table. If you use RMRIS to view, edit, or delete the sitelist you just created, and then use shift F2 (index) to bring up the sitelist, it will not be visible. You can still edit or delete it by simply typing in the name of the sitelist when the menu prompts you to do so. Sitelists loaded into the sitelist table are not visible by the index option through RMRIS. Therefore you must keep track of the sitelist files you create. For a complete listing of the sitelists stored in the sitelist table, use SQL to Query the sitelist table by typing:

```
SELECT DISTINCT SITE_LIST_NAME FROM R2RIS_SITE_LISTS;
```

You should make as many queries of the sitelist as needed to assure that the sites in each sitelist belong there and have accurate information. For example, a sawtimber site which was recently clearcut and is now a nonstocked site may not have had the tree data eliminated from the data base. A query on tree\_size will correctly

show that the site is nonstocked, but will still have complete tree data from the pre-cut measurement.

7. Use the sitelist to extract sample stand data to represent that stratum.
8. Plot the selected stands on a map using GIS to check their spatial distribution.

## REFERENCES CITED

- Cochran, William G. 1963. Sampling techniques, 2nd Ed. John Wiley and Sons, New York. 413 p.
- Cochran, William G. 1977. Sampling techniques, 3rd Ed. John Wiley and Sons, New York. 428 p.
- Dixon, Wilfrid J.; Massey, Frank J. Jr. 1969. Introduction to statistical analysis, 3rd Ed. McGraw-Hill Book Co. New York. 638 p.
- Freese, Frank. 1962. Elementary forest sampling. USDA Forest Service, Ag. Handbook No. 232. U.S. Dept. Agriculture, Washington, D.C. 91 p.
- Freese, Frank. 1967. Elementary statistical methods for foresters. USDA Forest Service, Ag. Handbook No. 317. U.S. Dept. Agriculture, Washington, D.C. 87 p.
- Grosenbaugh, L. R. 1965. Three-pee sampling theory and program 'THRP' for computer generation of selection criteria. USDA Forest Service Research Paper PSW-21. Pacific Southwest Forest and Range Experiment Station, Berkeley, CA. 53 p.
- Iverson, David C.; Alston, Richard C. 1986. The genesis of Forplan: a historical and analytical review of Forest Service planning models. Gen. Tech. Rept. INT-214. Ogden, UT: USDA Forest Service, Intermountain Forest and Range Experiment Station. 31 p.
- Lund, H. Gyde. 1975. Probability proportional to prediction (3P) sampling: an annotated bibliography. Unnumbered miscellaneous publication. USDA Forest Service, Northeastern Area, State and Private Forestry, Resource Use and Management Unit. Upper Darby, PA (Hq currently at Radnor, PA). 25 p.
- Lund, H. Gyde. 1978. Type maps, stratified sampling and PPS. Resource Note BLM 15. USDI Bureau of Land Management, Denver, CO. 18 p.
- Lund, H Gyde; Thomas, Charles E. 1989. A primer on stand and forest inventory designs. Gen. Tech. Rep. WO-54. Washington, DC: USDA Forest Service. 96 p.
- Mendenhall, William; Ott, Lyman; Scheaffer, Richard L. 1971. Elementary survey sampling. Wadsworth Publishing Company, Inc., Belmont, CA. 247 p.
- Sokal, Robert R.; Rohlf, F. James. 1969. Biometry. San Francisco: W. H. Freeman and Co. 776 p.
- Stage, Albert R. 1971. Sampling with probability proportional to size from a sorted list. Res. Pap. INT-88. Ogden, UT: USDA Forest Service, Intermountain Forest and Range Experiment Station. 16 p.
- Stage, Albert R.; Alley, Jack R. 1972. An inventory design using stand examinations for planning and programming timber management. Res. Pap. INT-126. Ogden, UT: USDA Forest Service, Intermountain Forest and Range Experiment Station. 17 p.

- USDA Forest Service. 1990. Resource inventory handbook. Forest Service Handbook 1909.14.
- USDA Forest Service. 1992. Timber resource planning handbook. Forest Service Handbook 2409.13.

## APPENDIX

The following is distributed with a Fortran executable program which is part of the routine for stand selection using the RIS data base. This material is available for retrieval from the Region Two DG computer R02A and is contained in a DGAOS "dumpfile". When executing the retrieval process, use this location address:

Host:	R02A
Staff:	RR_STAFF
Drawer:	LIBRARY
Folder:	DUMPFILES
Object:	RMRIS_SELECT_STANDS.DMP

The dumpfile contains five files as of this writing (April 1995):

SELECT\_STANDS.README  
SELECT\_STANDS.FOR  
SELECT\_LP9.SQL  
SELECT\_LP9.LIS

and the executable file of SELECT\_STANDS.FOR, named:

SELECT\_STANDS.PR

### Text of File SELECT\_STANDS.README

A valid set of sample stands may be selected which represent forested non-wilderness lands. The validity of this selection depends on several factors, including:

- 1) How much stand examination data exist on your Forest?
- 2) How well are those data distributed across Districts?
- 3) Were entire locations sampled, or just high volume stands?
- 4) The history of stand examination data collection on your Forest.
- 5) The validity of your area map (RIS DB) by Forest Type and Tree Size Class.

For Forest Inventory purposes, the idea is to select samples that are well-distributed and which have a probability of selection proportional to stand area.

For maximum flexibility and to have a minimum of 5 samples selected within any single stratum, samples are selected within Forest Type and Tree Size categories. The number of samples specified depends on how much further stratification by type and size the Forest intends to do for the purpose of Land Management Planning. For statistical analysis, all stand samples selected within a Forest Type and a Tree Size stratum have equal weight, so that simple averages may be applied.

Before selecting or creating your own data set, check with the Regional Office. A valid data set of either permanent plots or stand samples may already have been compiled based on the latest Forest Inventory. In the past, inventories usually collected around 300 samples to represent a Forest. Experience has shown us that between 600 and 1000 samples are more desirable for Forest Planning needs.

The best way to sample stands proportional to area and also to meet distribution concerns is to place a grid over the forest. With a grid, larger samples have a greater chance of selection and the nature of the grid insures distribution. However, you will find that some stands do not have data available, so you must either collect data for those stands or choose a denser grid.

Another way to sample stands proportional to area is to use a simple program designed to do this. Distribution across Districts may be assured by selecting separately for each District. Distribution within a District will be best if sites are ordered within locations. The number of samples on a District for a stratum must be proportional to the area of that stratum on the District.

The following is documentation for the sample selection program SELECT\_STANDS.PR used with the RMRIS data base for Forested non-wilderness:

1. list proc forest, location, site, and sitelist name  
xx xxxxxxx,xxxx  
in the above format into a file using ORACLE  
select  
set linesize 40  
spool select\_lp9  
select proc\_forest,location,site,area from r2ris\_site  
where  
cover\_type = 'TLP' and tree\_size = '9'  
and  
(special\_unit != '4' or special\_unit is null)  
order by proc\_forest,location,site;  
spool off

See sample SELECT\_LP9.SQL  
See sample output SELECT\_LP9.LIS

- 2 Edit the output file using sed/no\_form\_feed select\_lp9.lis, and remove the last three useless lines generated by the selection program.
3. Run this program and specify select\_lp9.lis (or whatever you called the output file from the previous procedure) as the input file.
4. Specify a desired AREA interval for the selection proportional to stand area — i.e., enter 3000 to select one stand sample for every 3000 acres.
5. Load the resulting output in the file STRATA into the R2RIS\_SITE\_LISTS table of RIS ORACLE via the RMRIS\_LOAD\_SITELIST.CTL macro using the command:

```
SQLLOAD /  
control file = RMRIS_LOAD_SITELIST.ctl
```

6. Use the SITE LIST with a joined query to determine how many of the selected stands have INTENSIVE stand examination data available.

```
Select proc_forest location, site  
from rmstand_header_data  
where survey_method = 'I' and  
(proc_forest,location,site) in  
(select proc_forest,location,site  
from r2ris_site_lists where site_list_name  
= 'LP9XX')
```

7. Narrow the site list down to the stands which have data available.
8. If there are not enough stands left for a sample stratum, begin the selection process again using a smaller acre interval, but do not bias the results by using more samples per stratum area on any one District.
9. Use the site list to extract sample stand data to represent that stratum.
10. Plot the selected stands on a map using GIS to check their spatial distribution.

## Text of File SELECT\_STANDS.FOR

```
c ----- Program to select stand samples proportional to area
c
c 1. list proc forest, location, site, and sitelist name
c xx xxxxxxx,xxxx
c in the above format onto a file using ORACLE select
c set linesize 40
c select proc_forest,location,site,area from r2ris_site
c where
c cover_type = 'TLP' and tree_size = '9'
c and
c (special_unit != '4' or special_unit is null)
c order by proc_forest,location,site;
c spool off
c
c See Sample SELECT_LP9.SQL
c See Sample Output SELECT_LP9.lis
c
c 2. sed/no_form_feed select_lp9.lis and remove the last
c three garbage lines
c
c 3. Run this program and specify select_lp9.lis or whatever
c you called it as the input file
c
c 4. Specify a desired AREA interval for the selection
c proportional to stand area — i.e. 3000 will select
c one stand sample every 3000 acres
c
c 5. Load the resulting output on the file STRATA into
c the R2RIS_SITE_LISTS table of RIS ORACLE via the
c RMRIS_LOAD_SITELIST.CTL macro using the command
c
c SQLLOAD /
c control file = RMRIS_LOAD_SITELIST.ctl
c
c 6. Use the SITE LIST with a joined query to determine
c how many of the select stands have INTENSIVE stand
c exam data
c
c Select proc_forest location,site
c from rmstand_header_data
c where survey_method = 'I' and
c (proc_forest,location,site) in
c (select proc_forest,location,site
c from r2ris_site_lists where site_list_name
c = 'LP9XX')
c
c 7. Narrow the site list down to the stands with data
c
c 8. If not enough for a stratum, then restart with
c a smaller acre interval
c
c 9. Use the site list to extract sample stand data
c to represent that strata
c
c 10. Plot out the selected stands on a map with GIS to
c check out the distribution
c
c character*5 ifile
```

```

c      icount = 0
c      write(*,fmt='(1h1)')
c
c      print *,''
c      print *,' PROGRAM SELECT STANDS REVISION 1.0'
c      print *,''
c      print *,'Reads list of Proc_Forest, Location and Site and Se-
c      lects'
c      print *,' Stand Samples based on Area onto a file named'
c      print *,''
c      print *,' STRATA.DAT'
c      print *,''
c      print *,' Load this file into a RIS Sitelist'
c      print *,' via the rmris_load_sitelist.ctl file'
c      print *,''
c
c      1 print *,''
c      write(*,fmt='(" Enter a 5 DIGIT File Name UPPER CASE=
c      ? ")')
c      read(*,err=1,fmt='(a5)')ifile
c
c      2 print *,''
c      write(*,fmt='(" Enter the Area Selection Interval = ? ")')
c      read(*,*,err=2)inc
c
c      iyes1= 'Y'
c      iyes2= 'y'
c
c      print *,''
c      write(*,fmt='(" Continue ? ")')
c      read(*,fmt='(a1)')icont
c      if (icont.ne.iyes1.and.icont.ne.iyes2) stop
c
c      write(*,fmt='(1h1,/," SELECTION RESULTS . . . ",/)')
c
c      write(*,fmt='(/," Reading File ",a5)')ifile
c      write(*,fmt='(/," Area Selection Interval = ",I10)')inc
c      print *,''
c
c      open (unit=1,file=ifile,err=1,blank='zero',form='formatted',
c      +status='old',maxrecl=132,pad='yes',recfm='ds')
c      open (2,file='strata.dat')
c
c      itotal = 0
c      10 read(1,fmt='(i2,1x,i6,1x,i4,1x,i10)',end=20)ifor,
c      iloc,isite,iarea
c      itotal = itotal + iarea
c      if (itotal.ge.inc) then
c          icount = icount + 1
c          write (*,fmt='(1h+,"Number Selected = ",i6)')icount
c          write (2,fmt='(i2.2,"",i6.6,"",i4.4,"",a5)')ifor,+
c          iloc,isite,ifile
c          itotal = 0
c      endif
c      go to 10
c      20 print *,''
c      print *,'END of SELECTION - Selected Sites are on file
c      STRATA.DAT'
c      end

```

### **Text of File SELECT\_LP9.SQL**

```
spool select_lp9
set pagesize 0
set linesize 40
select proc_forest,location,site,area from r2ris_site
where
cover_type = 'TLP' and tree_size = '9'
and
(special_unit != '4' or special_unit is null)
order by proc_forest,location,site;
spool off
```

### **Text of File SELECT\_LP9.LIS**

Note: Variable definitions and column headings added here.

Data variables are:

Columns 01-02 Proclaimed Forest Number  
Columns 04-09 Location-Compartment Number  
Columns 11-14 Location-Site Number  
Columns 16-25 Site Area-Acres

Column Numbers:

00000000011111111222222  
1234567890123456789012345

---

01 162103 0002 81  
01 162103 0003 49  
01 162103 0005 54  
01 162103 0006 34  
01 162103 0007 33  
01 162103 0008 21  
01 162103 0009 43  
01 162103 0013 56





The United States Department of Agriculture (USDA) prohibits discrimination in its programs on the basis of race, color, national origin, sex, religion, age, disability, political beliefs and marital or familial status. (Not all prohibited bases apply to all programs.) Persons with disabilities who require alternative means for communication of program information (braille, large print, audiotape, etc.) should contact the USDA Office of Communications at (202) 720-5881 (voice) or (202) 720-7808 (TDD).

To file a complaint, write the Secretary of Agriculture, U.S. Department of Agriculture, Washington, D.C. 20250, or call (202) 720-7327 (voice) or (202) 720-1127 (TDD). USDA is an equal employment opportunity employer.





1022409369



Rocky  
Mountains



Southwest



Great  
Plains

U.S. Department of Agriculture  
Forest Service

## Rocky Mountain Forest and Range Experiment Station

The Rocky Mountain Station is one of eight regional experiment stations, plus the Forest Products Laboratory and the Washington Office Staff, that make up the Forest Service research organization.

### RESEARCH FOCUS

Research programs at the Rocky Mountain Station are coordinated with area universities and with other institutions. Many studies are conducted on a cooperative basis to accelerate solutions to problems involving range, water, wildlife and fish habitat, human and community development, timber, recreation, protection, and multiresource evaluation.

### RESEARCH LOCATIONS

Research Work Units of the Rocky Mountain Station are operated in cooperation with universities in the following cities:

Albuquerque, New Mexico  
Flagstaff, Arizona  
Fort Collins, Colorado\*  
Laramie, Wyoming  
Lincoln, Nebraska  
Rapid City, South Dakota

\*Station Headquarters: 240 W. Prospect Rd., Fort Collins, CO 80526